

<https://helda.helsinki.fi>

Tekoäly ja semanttinen web tarjoavat huikeat mahdollisuudet suurten aineistojen hallintaan

Hyvönen, Eero

Valtion taidemuseo

2020-05-01

Hyvönen , E 2020 , Tekoäly ja semanttinen web tarjoavat huikeat mahdollisuudet suurten aineistojen hallintaan . julkaisussa J Nuorteva & P Happonen (toim) , Kansallisarkiston strategia 2025 : näkökulmia tulevaan . Kansallisarkiston toimituksia , Valtion taidemuseo , Helsinki , Sivut 20-21 . < https://arkisto.fi/uploads/Julkaisut/muut/KA_Strategia_2025.pdf >

<http://hdl.handle.net/10138/320364>

unspecified

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Tekoäly ja semanttinen web tarjoavat huikeat mahdollisuudet suurten aineistojen hallintaan

EERO HYVÖNEN

Aalto-yliopiston mediateknologian professori,
Helsingin yliopiston HELDIG:in (Helsinki Centre for
Digital Humanities) johtaja



STRATEGINEN TAVOITE 3.1.

Kansallisarkisto tekee aktiivisesti kansainvälistä ja kansallista yhteistyötä yliopistojen ja muiden toimijoiden kanssa digitaalisen tutkimusprosessin kehittämisessä.

Yliopistot ja Kansallisarkisto ovat tehneet hyvää yhteistyötä, mutta sitä voisi kehittää vielä systemaattisemmaksi. Yksi mahdollisuus olisi hyödyntää yhteistyön laajentamisessa HELDIG:in (Helsinki Center for Digital Humanities) hackathoneja – asetettujen kysymysten yhteiseen ja luovaan ratkaisemiseen tähtääviä käyttäjäryhmiä. Lupaava polku yhteistyölle ovat kansainväliset isot hankkeet ja EU-ohjelmat, kuten vireillä oleva Time Machine -hanke, jonka valmistelussa Kansallisarkisto on jo aktiivisesti mukana. Tällaisiin laajempiin kokonaisuuksiin pitäisi vahvemmin kytkeä myös kansallisia rahoittajia, kuten Suomen Akatemia.

Aalto-yliopisto ja Kansallisarkisto ovat toteuttaneet sotahistoriallisiin aineistoihin perustuvat linkitetyn datan palvelut ja semantti-

Professori Eero Hyvönen. Kuva: Eero Aalto.



Professori Eero Hyvösen työryhmän kehittämä semanttisen webin sovellus Sotasampo on saanut suuren suosion. Kuva: Kansallisarkisto.

set portaalit Sotasampo ja Sotasurmasampo 1914–1922. Nämä helpottavat esimerkiksi historioitsijoiden työtä, sillä he voivat hyödyntää ja analysoida niissä olevaa dataa ilman ohjelmointitaitoja. Vastaavantyyppisiä tutkimusinfrastruktuureita kehitetään jatkossakin.

Aineistot ja palvelut siirtyvät verkkoon, jonka teknologia kehittyy nopeasti yhä semanttisemmaksi ja älykkäämmäksi. Jääkiekkotähti Wayne Gretzky'n mukaan ei pidä luistella sinne, missä kiekko on nyt, vaan sinne, minne se on seuraavaksi menossa. Tämä edellyttää panostamista tutkimukseen ja omaan kehitystyöhön, koska seuraavan polven järjestelmiä ei voi aina ostaa kaupasta. Esimerkkinä tästä on käsinkirjoitetun tekstin tunnistuspalvelun kehittänyt EU:n READ-projekti. Uusien palveluiden ideointi, kehittäminen ja hyö-

dyntäminen digitaalisten ihmistieteiden tutkimustyössä ovat alueita, joissa avautuu yliopistoille ja Kansallisarkistolle kiinnostavia yhteistyömahdollisuuksia.

Yhteistyö muiden muistiorganisaatioiden ja dataa tuottavien tahojen, kuten kansalaisten, välillä tulee myös yhä tärkeämmäksi, sillä kulttuurialan sisällöt linkittyvät voimakkaasti toisiinsa organisaatorajoista välittämättä. Aineistojen automaattisen yhteentoimivuuden edistämiseksi tarvitaan yhteisesti sovittuja sanastoja ja ontologioita. Suomessakin pitäisi käyttää kansainvälisiä verkkoaineistojen ja linkitetyn datan julkaisukäytäntöjä ja työkaluja. Tämä mahdollistaa aineistojen rikastamisen myös kansainvälisten yhteistyökumppaneiden kanssa.

Yliopistojen kannalta kiinnostavaa ovat tutkimukselliset haasteet tiedonhallinnan menetelmien kehittämisessä, mutta myös käytännön tutkimustiedon tiedonhallintaan liittyvät kysymykset. Yliopistoissa koulutetaan ammatillaisia muistiorganisaatioiden tarpeisiin, joten yhteistyömahdollisuuksia



eri alueilla on runsaasti. Kustannusten ja resurssisen säästämiseksi tarvittaisiin kansallisen tason ratkaisuja ja palveluita.

Tarpeelliset ja hyödyllisiksi osoittautuneet tutkimusinfrastruktuurit pitäisi saada Suomen Akatemian tutkimusinfrastruktuurien tiekartalle, mikä parantaisi niiden mahdollisuuksia pitempiaikaiseen kehittämis- ja ylläpitorahoitukseen. Tärkeää olisi myös saada infrastruktuurit kytkettyä osaksi EU-tason ERIC-infrastruktuuriohjelmia, kuten humanististen alojen DARIAH:a

(Digital Research Infrastructure for the Arts and Humanities). Helsingin yliopiston HELDIG-keskuksessa ja Aalto-yliopistossa tähän pyritäänkin aktiivisesti. Toivottavasti tähän saadaan mukaan myös yliopistojen ulkopuolisia tahoja, kuten arkistoja, kirjastoja ja museota, joiden rooli datan omistajina ja omien aihealueittensa asiantuntijoina on ratkaisevan tärkeä. Infrastruktuuriohjelmien kautta vakiintuvat infrastruktuurit siirtyvät vähitellen osaksi organisaatioiden normaaleja palveluita ja toimintaa ja budjettirahoituksen piiriin.

Kansallisarkisto osallistuu EU:n rahoittamaan kansainväliseen Time Machine-hankesuunnitteluun. Kuva: Kansallisarkisto.